Objective Assessment of Speech Distribution Quality by means of Unconventional PESQ Application – A Case Study

Jan Holub, Jana Motlová, Radislav Šmíd, Ondřej Tomíška

FEE CTU Prague, Dept. of Measurement K13138, Technická 2

CZ 166 27 Prague 6, Czech Republic

holubjan@fel.cvut.cz

Abstract

The article describes non-traditional deployment of ITU-T P.862 (PESQ) algorithm for quality assessment of speech distribution. Subjective and objective results are presented and compared.

1. Introduction

Quality of Speech Distribution (e.g. by loudspeaker system in hospital) is often assessed by one of well known methods of speech intelligibility testing [1], [2], [3]. Objective (=not based on human listeners) methods are typically based on Speech Transmission Index (STI) measurement or its modifications (RASTI).

Although STI-based methods can theoretically work on real speech signal, the measurement is usually performed using optimized non-speech test signals (e.g. tone sequence) to reduce test duration. If one uses real speech signal as input for STI-based methods, the amount of recorded data required to gather enough information for modulation index calculation in each frequency band with satisfying accuracy would be enormous. On the other hand, artificial test signals (like tone or noise sequences) are disturbing the tested place as well as its neighboring areas, thus being unacceptable in certain cases (e.g. testing of sound distribution system in operational hospitals).

2. PESQ

Perceptual Evaluation of Speech Quality [4] is an algorithm working with real speech samples and designed for transmission quality measurements [5], [6]. Its use for speech intelligibility testing is not recommended for many reasons. E.g. it is known that combination of impairments that includes also signal reflections (echo) is not evaluated by PESQ correctly, means proportionally to human perception [7].

3. Unconventional PESQ Deployment

It was shown several times that even for non-standard, non-traditional applications that are not explicitly listed in P.862 and/or related standards, PESQ can, sometimes after proper modifications/ enhancements, provide meaningful results [8]. Therefore and despite the above given facts, we have tested the PESQ deployment for speech distribution system testing for the following case:

- The targeted application is operational check of distributed speech quality in case where non speech signals or long records of speech signals can not be used
- The electrical chain as well as electro-acoustical device (high quality active studio monitor in our case) contribution to resulting quality decrease is negligible
- The main source of decreased speech quality are (multiple) signal reflections and reception of signals coming from multiple sources with various delays and attenuations
- Background noise in the tested environment does not exceed typical office levels and is definitively less than 50 dB SPL (A).

4. Test Methodology

Studio quality recordings were made by two male and two female speakers. These four phonetically balanced

sentences were played three times in sequence in given tested place.
The distributed speech was recorded by high quality microphone. In some tested places the distributed speech was recorded in different microphone orientations, see Fig. 1. The recordings were processed by PESQ.
The PESQ MOS values from all talker type were averaged over given condition. Mean values and standard deviations were calculated.

Four acoustically different environments have been tested:

- Acoustic chamber with excellent sound absorbing lining (reverberation time less than 190 ms, background noise under 13 dB SPL (A)) – designated as "lab"
- Small office ($50m^3$) with simple sound absorbing lining ("office")
- Large room ($500m^3$) without any acoustic measures ("hall")
- Long corridor without any acoustic measures ("corridor")

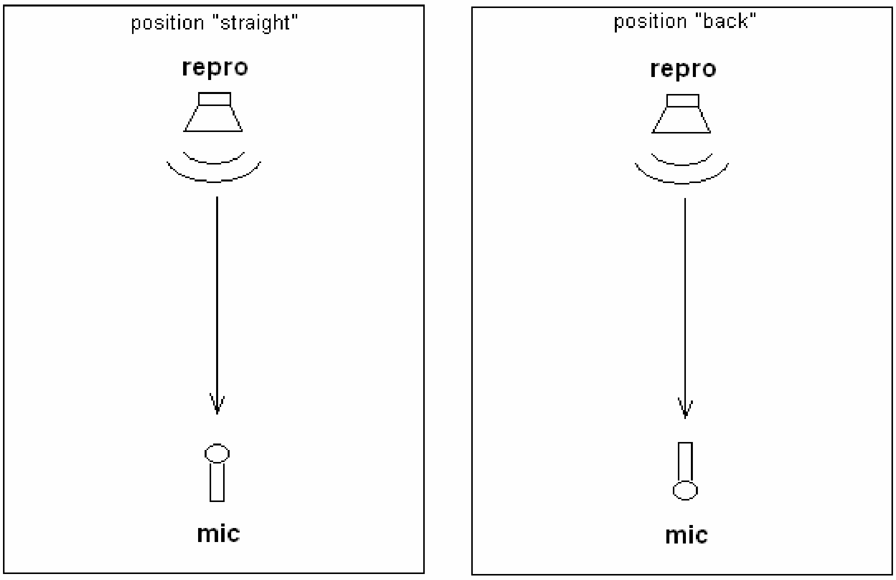

Figure 1: Microphone position

5. Results

Subjective listening tests have been performed on recorded distributed speech. Totally, more than 120 samples have been evaluated. The results of subjective testing have been averaged over given condition as well (chamber type and microphone position). These results are used as reference (etalon) values. MOS values obtained form PESQ were mapped into subjective results using 3-rd order regression. All three curves (1. PESQ MOS, 2. subjective results, 3. final estimated results from PESQ results) are shown on Fig. 2.

Environment types from 1 to 7 are as follows:

1. office_back
2. lab_back
3. lab_straight
4. hall_straight
5. hall_back
6. corridor_back
7. corridor_straight

Where "straight" means that microphone was oriented towards playback device and "back" means that microphone was oriented back side towards playback device (See Fig. 1).
It is obvious that although PESQ was not primarily designed for evaluating quality of distributed speech it can work pretty well under given conditions. The correlation coefficient between subjective and estimated quality assessment was 0.97.

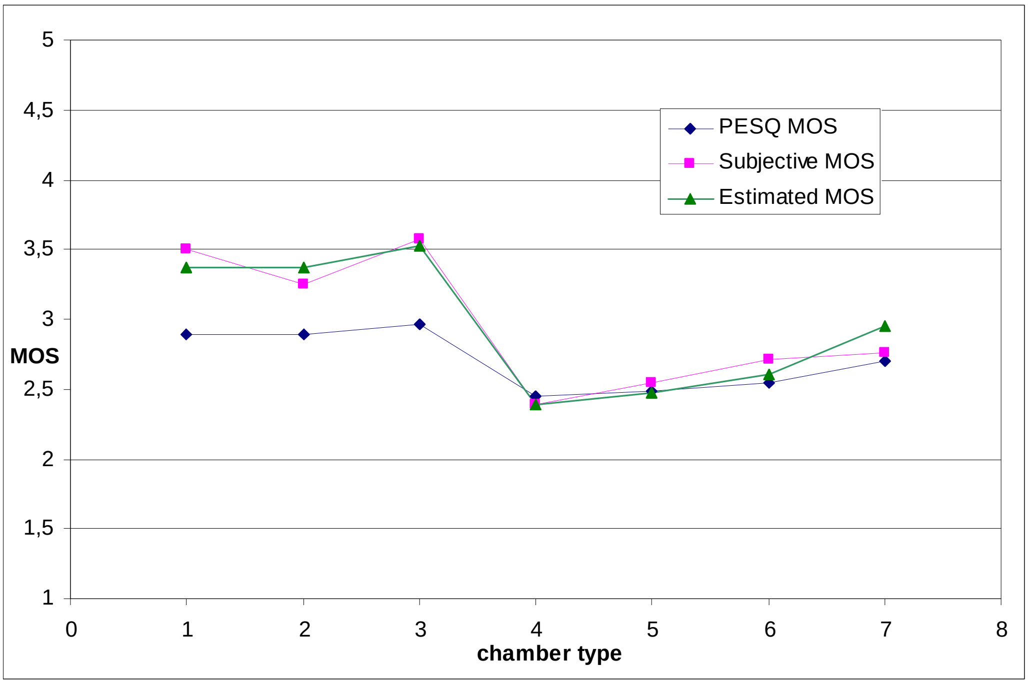

Figure 2: Results (the connecting lines have been used to increase the graph readability and have obviously no interpolating meaning

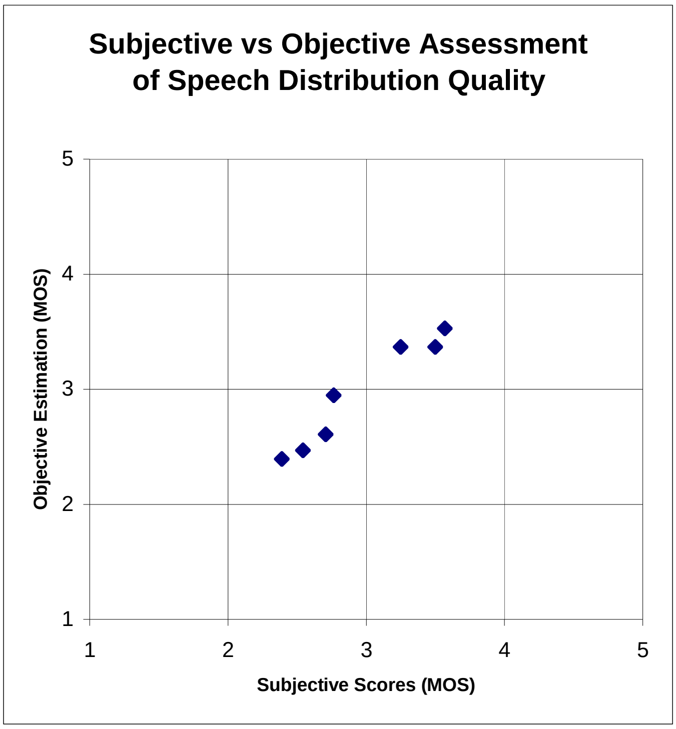

Figure 3: Results

6. Conclusion

Our current results show that under the above listed conditions, PESQ works quite satisfactorily as indicator of speech distributing system quality. This is not in contradiction with statements given in par. 3, as the criticized PESQ improper sensitivity to echoed signals is deployed in positive way in our application (PESQ is more sensitive to echo than human listener. In our approach, this contributes to increase measurement sensitivity and is finally removed by the 3-rd order regression at the test procedure end).

The experiment has shown a way how to objectively estimate distributed speech quality in situations where standard methods can not be used (e.g. in hospitals, standard doctor's messaging can be used as the test sentences).

Further work should be focused to wide-band PESQ (P.862.2) deployment, as the currently tested PESQ version works with narrow-band (speech) signal. Also more environment types and more records should be examined as current results span only within the range 2.3-3.6 MOS.

ACKNOWLEDGMENTS

This work has been supported by the Czech ministry of Education: MSM 6840770014 "Research in the Area of the Prospective Information and Navigation Technologies".

REFERENCES

[1] IEC 60268-16:2003: Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index

[2] ISO/TR 4870:1991: Acoustics - The construction and calibration of speech intelligibility tests

[3] Steeneken, H.J.M., Houtgast, T., "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am., vol. 67, pp. 318-326 (1980 Jan.)

[4] ITU-T P.862, Perceptual Evaluation of Speech Quality, ITU-T, February 2001

[5] Rix, A. W., Hollier, M. P, Hekstra, A. P., Beerends, J. G., "PESQ, the new ITU standard for objective measurement of perceived speech quality, Part 1 - Time alignment," J. Audio Eng. Soc., vol. 50, pp. 755-764 (2002 Oct.)

[6] Beerends, J. G., Hekstra, A. P., Rix, A. W., Hollier, M. P., "PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model," J. Audio Eng. Soc., vol. 50, pp. 765-778 (2002 Oct.)

[7] Holub, J., Šmíd, R., Bachtík, M.: Child Listeners as a Test Subject – Comparison with Adults and P.862, Measurement of Speech and Audio Quality in Network – MESAQIN 2002, June 2002, pp. 27-30, Prague

[8] Holub, J., Street, M., Šmíd, R.: Intrusive Speech Transmission Quality Measurements for Low Bit Rate Coded Audio Signals, AES115 Convention, New York, October 2003